DISTRIBUTED CRAWLING OF HYPERLINKED DOCUMENTS

5 **ABSTRACT**

Techniques for crawling hyperlinked documents are provided. Hyperlinked documents to be crawled are grouped by host and the host to be crawled next is selected according to a stall time of the host. The stall time can indicate the earliest

10 time that the host should be crawled and the stall times can be a predetermined amount of time, vary by host and be adjusted according to actual retrieval times from the host.